



Conception d'une Archive ouverte en SIC : le sens de la technique

Gabriel Gallezot, Christian Rossi, Ghislaine Chartron, Jean-Max Noyer

► To cite this version:

Gabriel Gallezot, Christian Rossi, Ghislaine Chartron, Jean-Max Noyer. Conception d'une Archive ouverte en SIC : le sens de la technique. Hypertextes hypermedias : créer du sens à l'ère du numérique, H2PTM'03, Sep 2003, Paris, France. pp.349-362. sic_00000943

HAL Id: sic_00000943

https://archivesic.ccsd.cnrs.fr/sic_00000943

Submitted on 15 Mar 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conception d'une Archive ouverte en SIC: le sens de la technique

Gallezot Gabriel, GRESI et LAMIC, Université de Nice, URFIST- Nice, Ave Joseph Vallot, 06108 Nice cedex 2, gallezot@unice.fr

Rossi Christian, CCSD/CNRS, Centre de Calcul de l'IN2P3, Villeurbanne, rossi@in2p3.fr

Chartron Ghislaine, GRESI, URFIST-Paris, 17 rue des Bernardins, 75005 Paris, chartron@cnam.fr

Noyer Jean-Max, CCI / CRECI, Université Paris 7, 2, place Jussieu, 75005 Paris., Jean-Max.Noyer@uhb.fr

RESUME

Cette communication présente la conception et la mise en œuvre d'une archive ouverte de publications scientifiques dans le champ des sciences de l'information et de la communication francophone. Dans un premier temps, il s'agit de comprendre le sens de la dialectique homme/machine associée au développement de ces archives ouvertes dans le cadre de la communication scientifique, la genèse du mouvement et les différentes orientations actuelles sont rappelées. Plus spécifiquement, les auteurs analysent les enjeux liés au développement de ce type de dispositif pour les sciences de l'information et de la communication dans l'espace francophone. Prenant en considération les spécificités de cette communauté, les auteurs expliquent ensuite leurs choix concernant l'architecture de cette archive, la conception de l'interface-usager, l'organisation et le fonctionnement de l'archive. Enfin les premières statistiques d'usages d'@rchiveSIC (<http://archivesic.ccsd.cnrs.fr>) sont exposées ainsi que les différents scénarios envisageables pour atteindre une masse critique suffisante donnant sens à cette innovation de services.

ABSTRACT :

This communication presents the design and the implementation of an Open Archive to scientific publications in french community of Communication and Information Sciences. First, understanding sens of dialectic Human/Machine associated the development of these Open Archive within the framework of the scientific communication, the genesis of the movement and the various current orientations are pointed out. More specifically, authors analyze the stakes related to the development of this type of device for the french community of Communication and Information Sciences. Taking into account specificities of this community, the authors explain then their choices relating to the architecture of this archive, the design of the user interfacier, the organization and use of the archive. Finally the first statistics of uses of @rchiveSIC (<http://archivesic.ccsd.cnrs.fr>) are exposed as well as the various possible scenarii to reach a sufficient critical mass giving sens to this innovation of services.

MOT-CLES : Open Archive, usage, information scientifique, communication scientifique

KEYWORDS : Open Archive, use, scientific information, scientific communication

1 - Les archives ouvertes dans le processus de communication: définition et enjeux

La relation auteur/public qui prévaut dans la communication scientifique relève encore du modèle classique établi par Garvey / Griffith dans les années 70¹ [CRA 96]. L'édition reste majoritairement sous forme papier et organisée par des éditeurs. Sans remettre en cause la fonction organisatrice et le travail des éditeurs, il faut aussi admettre que les connaissances inscrites sur les différents supports que représentent les revues ne circulent pas de façon optimum [HAR 01].

Les archives ouvertes de publications scientifiques sont l'exemple d'une négociation entre un processus social (la communication entre chercheurs) et un dispositif technique (la base de données accessible sur les réseaux), elles s'inscrivent dans la recherche d'une dialectique homme/machine, se distinguant d'une posture purement critique pour lui préférer une logique d'action conduisant à différentes expérimentations.

Quelques pionniers ont tracé de nouvelles voies en ce sens: Ginsparg a su valoriser la culture d'échanges de sa communauté en organisant une vaste banque de données internationale de diffusion des pré-publications devenue aujourd'hui l'une des principales sources d'information de son domaine [GIN 94], Stevan Harnad, cogniticien milita pour le principe du « Scholarly Skywriting », à savoir l'utilisation du réseau pour maximiser la diffusion des écrits scientifiques au-delà des supports existants [HAR 91].

Le mouvement engagé par Paul Ginsparg, a permis le développement d'un questionnement des modèles éditoriaux en place. En dévoilant certaines des potentialités techniques et économiques offertes par le numérique, Ginsparg a amené les chercheurs à s'interroger sur leur lien de dépendance vis-à-vis des éditeurs traditionnels.

Plusieurs idées fortes organisent alors ce mouvement. La première, conformément au projet des membres fondateurs du Web, consiste à exploiter les potentialités associatives dans la mise en commun des ressources. La seconde consiste à exploiter le faible coût de fabrication eu égard à la capacité de dissémination des oeuvres. La troisième consiste à rendre visible, aux yeux d'un plus grand nombre, les différents états, à différentes étapes, de la production d'un document, et de permettre ainsi un plus large éventail de la critique. Ce point est important car il intervient en amont de la stabilisation et de la légitimation finale et traditionnelle par les pairs dont le système de filtrage est largement dominé par les éditeurs et un nombre relativement réduit et stable d'évaluateurs. La quatrième est que l'accroissement des documents disponibles, des réseaux de liens amont et aval appelle le développement de nouvelles fonctions éditoriales permettant de travailler sur le processus d'écriture lui-même, ses conditions de production et ses conditions de citation. Filtrage, navigation, gestion des points de vue peuvent rendre possible une meilleure appréhension de la vie du document ou bien de la communauté des oeuvres comme

¹ Le modèle de Garvey/Griffith met en évidence les différentes étapes d'une communication scientifique, de la fin de l'initiation d'une recherche à ses résultats cités dans d'autres travaux scientifiques

incomplétude en procès de production pour reprendre ici la formule de A.N. Whitehead [WHI 95].

Le concept d'archives ouvertes de publications reste toutefois polymorphe. Le terme d'« archive » est essentiellement synonyme ici de réservoir d'information. La dimension traditionnelle associée aux « archives des archivistes », avec le souci majeur d'une préservation sur le long terme conduite par une politique institutionnelle, n'est pas ici l'enjeu premier même si cette dimension progresse dans le cadre récent des archives institutionnelles (*cf. infra*).

Concernant le qualificatif d'ouverture, il désigne ici un accès ouvert « sans barrière » aux écrits scientifiques, notamment sans modalité de paiement pour le lecteur-chercheur, renvoyant alors en terme économique à l'hypothèse de la publication scientifique comme bien public pour la société. Mais dans le contexte technologique, le qualificatif d'ouverture fait aussi explicitement référence à l'architecture technique de ces archives, il concerne alors l'utilisation d'un ensemble de protocoles techniques facilitant l'interopérabilité dans une organisation distribuée.

Ces différentes interprétations peuvent donc conduire parfois à des confusions dans les propos. Pour notre part, nous le considérerons dans notre projet dans sa dimension politique, à savoir un répertoire de publications en accès ouvert, respectant ou non les protocoles OAI².

Le développement des archives ouvertes se fonde sur un processus social d'auto-archivage des publications scientifiques par les auteurs qui, tout en continuant à soumettre leurs articles dans des revues ou colloques, déposent parallèlement des versions pré-publications (avant l'évaluation scientifique d'une revue) ou post-publications (après l'évaluation scientifique d'une revue) dans ces bases de données hypertextuelles, l'objectif est de maximiser leur visibilité à l'heure des réseaux. Cette pratique nécessite des négociations avec les revues en terme de cession des droits liés à la propriété intellectuelle, l'idée étant de parvenir à la satisfaction des auteurs tout en préservant un modèle économique viable pour les revues.

Une question essentielle peut alors être posée: quels acteurs peuvent légitimement assurer la coordination d'un tel processus ? C'est une question essentielle qui explique en grande partie l'évolution actuelle du mouvement des archives ouvertes. L'idée selon laquelle ces archives ne remettent pas en cause les revues et leurs cercles de validation scientifique mais constituent uniquement un mode supplémentaire de diffusion est un message difficile à faire comprendre, les résistances se sont ainsi avérées nombreuses ces dernières années³. Pour cette raison et certainement d'autres liées à la défense des « territoires » établis, on assiste plutôt aujourd'hui au développement d'archives institutionnelles, valorisant les

² Open Archives Initiatives, <http://www.openarchives.org>

³ Les archives disciplinaires comme arXiv, CogPrints, Wopec, CiteSeer sont des bases centralisées qui ont connu des croissances différentes, arXiv est certainement la plus fédératrice dans la mesure où elle se greffait sur des pratiques sociales d'échanges préexistantes mais Cogprints par exemple initiée par S. Harnad connaît depuis un certain nombre d'années un ralentissement majeur, elle ne rassemble que 1500 articles malgré l'énergie déployée par son initiateur.

productions locales d'une seule institution. Certaines universités prestigieuses ont déjà su mettre à profit cette nouvelle forme de valorisation (*cf.* par exemple, l'Université de Californie⁴, du MIT⁵). Au niveau international, les bibliothèques de recherche y voient désormais une nouvelle mission qu'elles doivent investir pour leur université [SPA 02], [JOH 02].

Le projet intellectuel d'une archive disciplinaire et d'une archive institutionnelle n'est toutefois plus le même: dans un cas, il s'agit de mieux faire circuler les résultats dans un champ scientifique en proposant des accès transversaux aux écrits de recherche (revues, actes de colloques, prépublications), dans l'autre cas, il s'agit plus de constituer une archive au sens traditionnel du terme, à savoir une mémoire des productions locales (articles, thèses, rapports, cours en ligne) stockées sous une forme numérique et associée à une nouvelle forme de valorisation.

2 - Quels développements pour ces archives ouvertes?

Le nombre des archives ouvertes est difficile à apprécier, il n'existe pas de déclaration obligatoire de ces nouveaux types de contenu. Toutefois, certains répertoires tentent de les recenser:

- 59 archives sont recensées sur le site de l'OAI (Open Archives Initiatives), il s'agit d'archives auto-déclarées par les concepteurs et respectant le protocole OAI (dimension technologique)⁶,
- 14 archives sont recensées sur le site de l'OAF (Open Archives Forum⁷), site fédérateur des initiatives principalement européennes. Un bilan des initiatives européennes a récemment été publié par cet organisme [DOB 03] ,
- 59 initiatives figurent dans le répertoire Open Directory Project (ODP)⁸,
- Enfin, selon le site GNU Eprints⁹, 66 archives utilisent le logiciel Eprints (de l'Université de Southampton), actuellement l'un des principaux logiciels pour la conception des archives ouvertes.

Ces estimations recouvrent en fait une grande disparité de contenus: ces archives peuvent concerner des articles, des thèses, des rapports, des pré-publications. De plus, elles sont soit institutionnelles, soit plus transversales, essayant d'atteindre une masse critique disciplinaire.

Quid des archives ouvertes en sciences de l'information et de la communication? En consultant les 4 répertoires cités précédemment, on identifie actuellement les sources suivantes, sans garantie d'exhaustivité:

⁴ <http://repositories.cdlib.org/escholarship>

⁵ <http://www.dspace.org/>

⁶ <http://www.openarchives.org>

⁷ <http://www.oaforum.org>

⁸ http://dmoz.org/Science/Publications/Archives/Free_Access_Online_Archives

⁹ <http://software.eprints.org>

| Nom de l'archive | Acteurs | Adresse | Contenus |
|--|---|--|--|
| Caltech Library System papers and Publication | Archive institutionnelle Université Caltech (USA) | http://caltechlib.library.caltech.edu (section Library & Information Science) | 17 articles ou communications |
| Digital Library of Information Science and Technology (DLIST) | Archive institutionnelle de l'Université d'Arizona (USA) | http://dlist.sir.arizona.edu | 65 articles ou communications |
| Documents in Information Science (DoIS) | Consortium d'Universités (UK, ES, IT, USA) | http://dois.mimas.ac.uk | 10042 articles, 3045 communications de conférences mais uniquement 6928 documents sont téléchargeables en texte intégral |
| EPrints.FRI | Archive de l'Université de Ljubljana, Informatique et sciences de l'information (Slovaquie) | http://eprints.fri.uni-lj.si | 25 articles ou communications |
| Formations Media Studies | University of Ulster (UK) | http://formations.org.uk/ | 56 articles ou communications |
| Prints in Library and Information Science (ELIS) | Projet disciplinaire soutenu par le Ministère espagnol de l'éducation et par un consortium italien de 9 Universités | http://eprints.rclis.org | 62 articles ou communications ou rapports |
| @rchiveSIC Archive en sciences de l'information et de la communication | Projet disciplinaire initié par 3 laboratoires de recherche français, projet soutenu par le CNRS (France) | http://archivesic.ccsd.cnrs.fr | 119 Articles ou communications de colloques. |

Archives ouvertes recensées en sciences de l'information et de la communication (mars 2003)

Sources: <http://www.openarchives.org>, <http://www.oaforum.org>,
http://dmoz.org/Science/Publications/Archives/Free_Access_Online_Archives,
<http://software.eprints.org>

Plusieurs remarques sur ce premier inventaire:

- Il faut noter le faible nombre de documents dans 5 archives sur 7 recensées. Une archive se distingue et recueille aujourd'hui une masse critique importante de documents (DoIS). Nous n'avons pas pu recueillir sur les sites les dates de création, ce qui aurait permis de mieux apprécier ces chiffres.
- La stratégie institutionnelle aboutit à de petites archives dans cette discipline, l'émiettement est un risque fort probable.
- Les sciences de la Communication sont très faiblement représentées dans cet échantillon (2 archives), la spécificité française est de rapprocher Communication et

Information par rapport aux autres projets dans lesquels les sciences de l'Information sont plus proches de l'Informatique.

- Enfin, l'absence de certains pays européens est à noter: l'Allemagne, les pays nordiques en particulier.

3 - Pourquoi une archive ouverte francophone en SIC ?

Une récente enquête du Ministère de la Recherche en France [Jeannin, 2002], faisait état d'une situation alarmante concernant les revues en sciences de l'Information et de la Communication (SIC). Sur une liste de 263 revues recensées en SIC au plan mondial, 17 seulement sont connues d'au moins un chercheur sur deux. Quand on passe au plan national, 12 seulement des 40 revues françaises citées en SIC sont connues (mais pas forcément lues) d'au moins un chercheur sur deux. Ces chiffres montrent une dispersion et une méconnaissance des revues censées structurer notre discipline. [BOU 02]

Pour donner aux articles la lisibilité qu'ils méritent, pour faciliter la recherche et l'accès au savoir contenu dans les articles, pour relever le défi de la lutte contre la « babelisation » de la discipline tout en conservant la diversité qui en fait sa richesse, nous avons proposé à une instance de représentation de la communauté « information-communication » française la création d'une Archive ouverte : @rchiveSIC¹⁰ [GAL 02]. Cette proposition s'inscrit dans une logique de recherche-action articulant analyse des enjeux et expérimentation d'un dispositif. Il ne s'agit ni d'expérimenter aveuglément, ni d'en rester à une posture distanciée mais de s'inscrire dans un processus de transformation de la réalité de la publication et de produire des connaissances concernant ces transformations [HUG 88]. La conception est guidée par l'usage, alliant développement technique et pertinence de son insertion sociale. Ce présent texte propose des éléments de réflexions pour une discussion sur les transformations engagées par ce type de dispositif.

L'expérimentation de cette archive ouverte d'articles scientifiques devrait permettre:

- Une plus grande visibilité des articles,
- Un accès facilité à des écrits actuellement trop dispersés,
- Un encouragement à des échanges pluridisciplinaires sur les objets de recherche,
- Une cartographie de notre communauté,
- Une mémoire collective concourant à la structuration de la discipline
- Une réflexion sur les nouvelles fonctionnalités éditoriales liées aux variations des pratiques cognitives en cours

De plus, deux enjeux spécifiques nous paraissent distinguer ce projet de celui de nombreuses autres archives ouvertes de publications scientifiques:

- Une approche francophone, avec la volonté d'atteindre une masse critique suffisante au niveau international. L'échelle institutionnelle pourtant repérée comme étant plus développée dans les projets actuels d'archives ouvertes nous a semblé ici

¹⁰ <http://archivesic.ccsd.cnrs.fr>

peu crédible car les départements information-communication sont de taille très modestes dans la plupart des universités françaises, contrairement à d'autres disciplines.

- Le fait de considérer cette archive comme un outil de structuration de notre communauté rassemblant des approches, des épistémologies fort différentes pour étudier les processus informationnels et communicationnels, spécificité peut-être uniquement française de réunir les sciences de l'Information et les sciences de la Communication dans les instances d'évaluation. Ce point distingue en grande partie la motivation d'un tel projet des autres projets recensés dans la partie précédente.

4 La conception d'@rchiveSIC

L'architecture technologique

Pour réaliser cette archive nous nous sommes appuyés sur les compétences techniques du CCSD (Centre pour la Communication Scientifique Directe) du CNRS qui maintient déjà des archives ouvertes dont « thèses_EN_ligne » et le point d'entrée français d'ArXiv.

L'archive est construite à partir du logiciel eprints.org¹¹ (version 1.1). Eprint.org fonctionne sous Unix (PC linux...) avec un serveur Apache, une base de donnée MySQL et le langage Perl. Il est développé et maintenu¹² au département d'électronique et d'informatique de l'Université de Southampton sous la forme d'un logiciel libre (licence GNU GPL¹³). Ce logiciel est gratuit, librement modifiable et déjà utilisé par plus d'une soixantaine de sites (mars 2003). Il est compatible OAI et à cet égard permet de rendre interopérable notre projet avec d'autres archives. Partant des spécificités de notre communauté scientifique, nous avons adapté Eprint.org et ajouté les fonctionnalités suivantes dans le cadre de la réalisation d'@rchiveSIC:

- Simplification du processus de dépôt en 3 écrans au lieu d'une dizaine,
- Possibilité de multi-dépôts, tout en conservant la « thématization » de ceux-ci,
- Ouverture des dépôts à des formats fortement utilisés (rtf, .doc, .zip), vérification des en-têtes des formats, générateur automatique de format HTML et PDF,
- Modération partagée,
- Recherche full-text,
- Version bilingue (la version 1.1 d'Eprint.org n'offre pas en standard cette possibilité),
- Statistique de dépôt.

L'hébergement de l'archive est effectué sur les serveurs du Centre de Calcul de l'IN2P3 (CCSD/CNRS¹⁴) qui bénéficient d'un accès direct au réseau RENATER.

Si nous avons choisi Eprint.org, seul outil bien distribué à la date de création d'@rchiveSIC, d'autres logiciels ou plates-formes techniques sont aujourd'hui

¹¹ <http://www.eprints.org/>

¹² Robert Tansley, Christopher Gutteridge et Mike Jewell sous la direction de Stevan Harnad

¹³ <http://www.gnu.org/>

¹⁴ <http://www.ccsd.cnrs.fr/>

disponibles et proposent des solutions « clés en main » d'auto-archivage, on relèvera notamment : DSpace du MIT¹⁵, CDSware du CERN¹⁶ et HAL du CCSD¹⁷.

Si les enjeux semblent évidents et la solution technique éprouvée, un outil technique ne s'impose pas, il se négocie collectivement. Ainsi pour emporter l'adhésion des enseignants, chercheurs et praticiens de notre communauté scientifique nous avons profité du contexte offert par le BOAI¹⁸ et de l'occasion offerte par un colloque de la SFSIC¹⁹ sur les revues francophones en SIC pour lancer @rchiveSIC.

Fonctionnement de l'archive

Enregistrement des auteurs : l'archive est ouverte uniquement à la communauté des enseignants et chercheurs francophones en SIC. Les déposants s'enregistrent et fournissent un minimum d'informations sur leur identité. L'enregistrement est automatique, mais une vérification a posteriori est possible lors du dépôt de documents.

Dépôt par l'auteur et déposant : pour des raisons juridiques, c'est l'auteur lui-même qui est incité à déposer ses textes. En effet, en France dans notre communauté scientifique peu de contrats d'édition ou des contrats insuffisamment précis sont signés entre l'auteur et une revue. Pour éviter les problèmes juridiques, nous réclamons que le dépôt soit sous la responsabilité de l'auteur. Néanmoins, une autre personne enregistrée peut, avec l'accord écrit de(s) auteur(s) poster un texte et être ainsi un déposant.

Trois statuts de documents : trois types de documents sont acceptés. Les statuts « working paper », « soumis à validation », « publié ou en cours de publication » permettront à chacun de publier ses résultats au moment voulu.

Modération partagée : pour commencer à dessiner des points scientifiques nodaux au sein de notre communauté, les déposants doivent classer leurs documents selon 20 thématiques (plusieurs choix sont possibles), ces thématiques ont été progressivement ajustées dans la phase de test. Chaque thématique a son modérateur. Les modérateurs de l'archive veillent uniquement à ce que les articles déposés correspondent à la vocation de l'archive, ils assurent un filtrage minimum mais non une validation scientifique du contenu. De plus cette modération partagée élimine toute velléité éditoriale d'un modérateur puisqu'un texte peut être classé dans plusieurs thématiques et donc être validé ou invalidé par un autre modérateur.

Interface bilingue et dépôt multilingue : bien que l'archive s'adresse principalement aux francophones, l'interface est volontairement bilingue pour permettre une recherche mondiale des textes contenus dans l'archive. De la même façon, la langue de dépôt n'est pas restreinte au seul français, mais ouvert à toutes les langues dans lesquelles un auteur de la communauté a pu communiquer (principalement : anglais, allemand, espagnol, italien).

¹⁵ Durable Digital Repository , <http://www.dspace.org/>

¹⁶ CERN Document Server Software, <http://cdsware.cern.ch/>

¹⁷ Hyper article en ligne, <http://hal.ccsd.cnrs.fr/>

¹⁸ Budapest Open Access Initiative, <http://www.soros.org/openaccess/>

¹⁹ Société Française des Sciences de l'Information et de la Communication

Abonnement à des thèmes : chaque utilisateur de l'archive peut s'abonner aux thématiques scientifiques qui servent au classement des textes et choisir la périodicité de l'alerte des nouveaux textes.

Format de dépôt et de visualisation : quatre formats couramment utilisés dans notre communauté scientifique ont été retenus : RTF, DOC, PDF, HTML et dans une moindre mesure PS. Selon, la taille du texte ou sa composition en plusieurs fichiers, l'auteur peut aussi déposer un fichier compressé au format Zip, gz ou tar.gz. Tous les formats de dépôt sont visibles. Les formats PDF et HTML sont automatiquement générés si le dépôt s'est effectué dans d'autres formats, assurant ainsi un minimum de lisibilité, voir de pérennité.

L'interface : notre communauté comprend des enseignants, chercheurs et praticiens de domaines très diversifiés et qui possèdent un niveau d'appropriation des NTIC différent. Aussi, nous avons voulu une interface de dépôt très simple permettant en quelques mots et quelques clics de déposer son texte. La première page est constituée d'une notice bibliographique très sommaire (auteur, titre, statut du document, date, résumé, mots-clés), la deuxième d'une série de 20 thématiques scientifiques, la troisième du choix du format de dépôt des textes, la quatrième concerne le chargement des fichiers et la cinquième la validation des informations soumises au serveur avec la mention juridique de la responsabilité de l'auteur/déposant. L'interface d'interrogation est à cette image. Possibilité de recherche en full-text, recherche en feuilletant par thématique et recherche par mots-clés (simple et avancée)²⁰.

Les métadonnées : elles sont présentes uniquement dans la notice bibliographique du dépôt²¹. Elles respectent le standard OAI, qui s'exprime principalement par un balisage XML des éléments du Dublin Core. Ainsi, en respectant ce standard les textes présents dans l'archive peuvent être interconnectés à d'autres bases compatibles OAI. D'autre part, l'ensemble des notices se trouve dans un répertoire accessible aux moteurs de recherche, les métadonnées OAI permettent globalement une bonne indexation²².

²⁰ Pour plus de détail sur l'interface : <http://archivesic.ccsd.cnrs.fr>

²¹ Une feuille de style (DTD XML) aurait pu affiner le référencement du texte, mais aurait considérablement bloqué la dynamique de dépôt

²² Prise en compte du *Dublin core* et du principe de *Link popularity* par le standard OAI

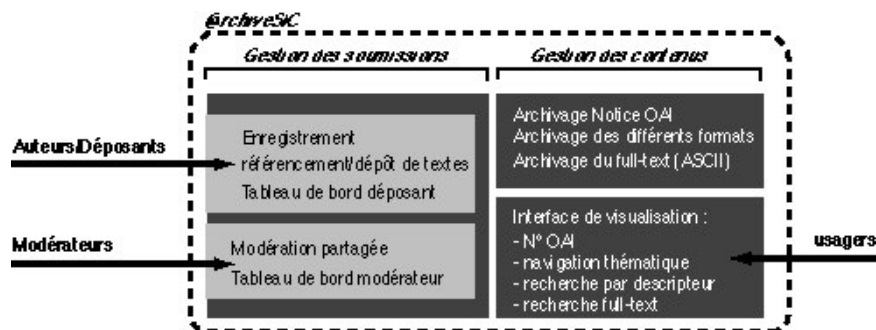
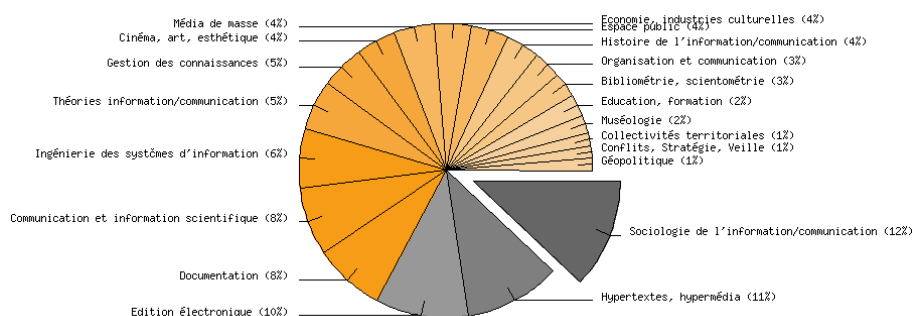


Schéma fonctionnel d'@rchiveSIC

5 – Premières statistiques et retours d'usage (mars 2003)

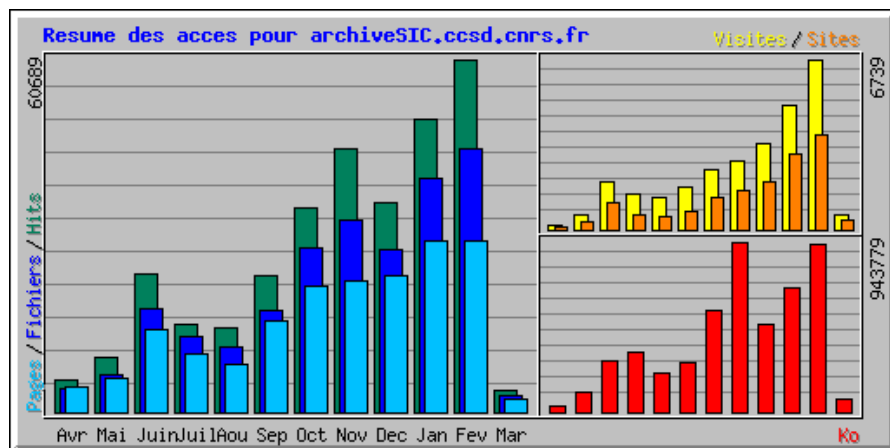
Statistiques²³

@rchiveSIC est disponible depuis juin 2002, elle compte 207 inscrits avec une progression d'inscription de 18 comptes par mois (déposants ou abonnement thématique). Elle contient à ce jour 119 documents, avec en moyenne 11 documents supplémentaires par mois. Ces documents représentent 293 entrées dans les différentes thématiques proposées et se répartissent selon le graphe suivant :



Concernant la consultation d'@rchiveSIC, le serveur totalise plus de 27 000 visites, avec une pointe de 6739 visites (qui émanent de 3741 sites différents) pour le mois de février 2003. Le détail du nombre de pages (pages vues), de fichiers (téléchargés) et de hits (accès) est présenté ci-dessous. Il faut aussi noter le passage très régulier des spiders des moteurs de recherche (google, fast, ...).

²³ Consultable à : <http://archivesic.ccsd.cnrs.fr/info/>



(Généré par Webalizer version 2.01)

Statistiques de consultation d'@rchiveSIC (mars 2003)

Retours d'usage

De façon informelle, nous avons aussi relevé les premières réticences au dépôt: qui s'occupe de l'archive ? Pourquoi le CNRS ? C'est quoi le CCSD ? Quelle politique éditoriale ? Finalité recherchée ? Que pensent les éditeurs de revues en info/com ? Quel statut pour les textes déposés ? ... Mais le frein principal au dépôt semble relever du rapport « temps/utilité ». Des déposants accaparés par d'autres charges de travail ne trouvent, actuellement pas d'utilité immédiate à ce projet.

6 - Perspectives de développement et interrogations

Les difficultés majeures à surmonter désormais ne sont pas d'ordre technologique mais d'ordre culturel. Comment étendre l'usage de l'archive dans les pratiques de nos collègues? Plusieurs stratégies sont actuellement envisagées en partant de l'hypothèse que l'appropriation doit permettre un gain d'efficacité pour chacun.

L'archive comme support de diffusion des actes de colloques et conférences

Il nous semble déceler à travers les dépôts déjà effectués et des engagements déjà inscrits dans des appels à communication une dynamique de dépôt liée à la publication de communication de colloques et de conférences (par exemple, le colloque « Les défis de la publication sur le Web Hyperlectures, cybertextes et méta-éditions » 9-11 décembre 2002 à Lyon, et le colloque CIFSIC, Bucarest 28 juin 2003). En effet, la publication d'actes est souvent coûteuse, généralement supportée par l'inscription au colloque et circule très mal compte tenu du nombre d'exemplaires édités. Pourtant, les savoirs, les idées, les résultats diffusés lors de ces manifestations sont généralement en émergence et constituent un substrat scientifique fertile. Aussi, un travail de recensement des colloques et des conférences est à réaliser et sera suivi d'une demande de partenariat avec @rchiveSIC.

L'archive comme mémoire des publications des équipes de recherche

Le principal frein à l'accroissement du nombre de textes dans l'archive reste le « manque de temps » des déposants. En effet, l'auteur doit souvent rechercher son papier sur son disque dur (voir sur ses disquettes ou encore le numériser), puis l'indexer, le charger et ce pour chacun de ses textes. Pour palier à cette « perte de temps », un bon moyen serait d'apporter un service supplémentaire pour la même action (le dépôt).

Aussi, nous pensons que la mise en place de « micro-archives » au sein de chaque équipe de recherche (ou autre entité) permettrait d'une part de constituer une mémoire des publications d'un laboratoire, d'un groupe et d'autre part d'alimenter par un principe de cross-posting la base @rchiveSIC. Chaque chercheur déciderait, bien entendu du niveau de diffusion de son article (interne à l'équipe ou public)²⁴.

L'archive comme diffusion parallèle des revues

A la lumière de quelques mois d'expérimentation d'@rchiveSIC, une autre dynamique de dépôt pourrait être réalisée. Il s'agirait de négocier avec les revues du domaine un temps de latence entre la publication papier et la publication électronique sur @rchiveSIC. Ce principe est déjà mis en œuvre et accepté par un nombre important de revues dans le domaine biomédical²⁵ avec souvent un délai de 6 mois. La version électronique peut ne pas être le fac-similé de l'édition papier, mais la version mise en forme par l'auteur sur son traitement de texte. Ainsi, l'évaluation scientifique et le travail d'édition restent inchangés, seule la diffusion des savoirs est démultipliée. Le processus inverse peut présenter un intérêt pour les éditeurs: sur le modèle d'ArXiv, un ensemble de pré-publications ou de working-papers déposés dans @rchiveSIC peuvent révéler une collection à retravailler pour constituer un objet éditorial à valeur ajoutée.

L'archive pour thématiser « la » discipline

Un questionnement concerne les thématiques choisies à la construction de l'archive pour « thématiser » les dépôts²⁶. Cette proposition de classification, forcément imparfaite et réductrice, est nécessaire pour mettre en place une modération partagée (cf. supra) et ne pas faire de l'archive un fourre-tout. Néanmoins, pour reprendre l'idée forte de la conception de l'archive, il s'agit de dépasser ce stade en évoluant vers une cartographie dynamique à partir de l'analyse du texte intégral. Si les catégories ne reflètent pas de façon exhaustive l'ensemble de la discipline info/com, l'ajout de catégories reste possible, mais elles se multiplieront, se subdiviseront, se recomposeront avec l'évolution de « la » discipline, enfermant les recherches dans des tiroirs mal dimensionnés. Pour visualiser les thèmes forts qui structurent la discipline, nous proposons de générer des cartes sémantiques (avec des méthodes

²⁴ Ce type de processus est d'ailleurs le fondement de DSpace

²⁵ freemedicaljournals : <http://www.freemedicaljournals.com/> et Pubmed : <http://www.pubmedcentral.nih.gov/>

²⁶ http://archivesic.ccsd.cnrs.fr/view_ROOT_fr.html

statistiques) qui permettraient un instantané thématique par lequel la navigation serait plus en adéquation avec le contenu évolutif de l'archive²⁷.

L'archive comme adjuvant aux articles édités

Si pour l'instant l'archive est utilisée comme une réplique électronique de la production papier, des appendices peuvent et devraient agrémenter les dépôts (données, images, annexes, films, liens, etc.). Le support papier, pour des raisons techniques ou de coût, ne se prête pas à l'ajout de documents connexes à l'article publié. Sur le modèle de la biologie moléculaire où la longueur des descriptions des séquences génétiques a nécessité la création de base de données ad hoc [GAL 00], @rchiveSIC peut servir d'entrepôt d'articles augmentés. De même, des services de liens de co-citations entre les déposants sont matérialisables²⁸.

L'archive comme aide à la construction des connaissances

Ce type de dispositif met en avant une nouvelle médiation entre l'auteur et son public. Au-delà de la disponibilité immédiate des textes (visualisation écran ou impression papier), il ne s'agit plus seulement de repérer un article à l'aide de mots-clés et de mettre ainsi en relation un auteur et un lecteur, mais de réaliser une exploration informationnelle d'un corpus de documents. Ce type de démarche peut favoriser la créativité des chercheurs et, par voie de conséquence, la construction de nouvelles connaissances. En fouillant les sédiments cognitifs accumulés dans l'archive, les résultats proposés mettent en lumière certains liens qui n'auraient pu être perçus autrement et qui peuvent faire sens auprès du chercheur [GAL 02].

Bibliographie

[BOU 01] Bounoux D., Rasse P., « Les revues en SIC », La lettre d'Inforcom, n°59, juin 2001.

[CHA 03] Charnay D., « Le libre accès : quels enjeux pour la communauté scientifique ? », Libre accès à l'information scientifique et technique : État de l'art et perspectives, Paris, 23-24 janvier 2003 - Carré des Sciences - Ministère de la Recherche, <http://www.inist.fr/openaccess/transparentes/charnay/index_fichiers/frame.htm>.

[CRA 96] Crawford S. Y., Hurd J. M., Weller A.C., From Print to Electronic, The Transformation of Scientific Communication, Medford, Information today Inc (Asis monograph series), Washington, USA, 1996.

[DEL 03] De la Vega J. F.), La communication scientifique à l'épreuve de l'Internet, l'émergence d'un nouveau modèle, Presses de l'Enssib, 2000.

[DOB 03] Dobratz S., Matthaei B., «Open Archives Activities and Experiences in Europe, An overview by the Open Archives Forum», D-Lib Magazine, Vol.9, n°1, janvier 2003, <<http://www.dlib.org/dlib/january03/dobratz/01dobratz.html>>.

²⁷ Lelu Alain, « Maquette de cartographie pour les 7 premiers numéros de la revue Solaris », ADEST, 13/11/2001, <http://hdyh.hymedia.univ-paris8.fr/neuroweb/solaris/cartoweb.html>

²⁸ Voir à ce sujet CiteSeer <http://www.neci.nec.com/~lawrence/aci.html> et OpCit <http://opcit.eprints.org/>

- [GAL 02] Gallezot G., Chartron G., Noyer J.-M., « Une archive ouverte des publications en InfoCom », colloque Place et enjeux des revues pour la recherche en infoCom , SFSIC, Nice, 25-26 Mars 2002,
<http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/00/25/index_fr.html>.
- [GAL 00] Gallezot G., Samson F., Brunaud V., Gas S., Bessière P. , « Normes et standards dans le processus de traitement du document numérique en biologie moléculaire », Solaris, n°6, 2000, <<http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/6gallezo.html>>.
- [GAL 02] Gallezot G., « Exploration informationnelle et construction des connaissances en génomique », Les Cahiers du numérique, Hermès, vol.3, n°3, novembre 2002.
- [GIN 94] Ginsparg P., “First Steps Towards Electronic Research Communication”, Computers in Physics, vol.8, n°4, 1994, repris dans Solaris n°3,
<<http://www.info.unicaen.fr/bnum/jelec/Solaris/d03/3ginspar.html>>.
- [HAR 91] Harnad S., “Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge”, Public Access Computer Systems Review 2 (1), 1991, pp.39-53. Voir également l’ensemble des publications de Stevan Harnad à l’adresse
<<http://cogsci.soton.ac.uk/~harnad/intpub.html>>.
- [HAR 01] Harnad S., “For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving”, 2001,
<<http://cogprints.soton.ac.uk/documents/disk0/00/00/16/39/index.html>>.
- [HUG 88] Hugon Marie-Anne, Seigel Claude, « Recherches impliquées, Recherche-action : le cas de l’éducation », Belgique Deboek université, 1988.
- [JEA 01] Jeannin P., Enquête sur les périodiques en Sciences de l’Information et de la Communication, 2001, <http://www.iut-tarbes.fr/enquete/info_comm/>.
- [JOH 02] Johnson R. K., “Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication”, D-Lib Magazine, Nov 2002,
<<http://www.dlib.org/dlib/november02/johnson/11johnson.html>>.
- [SPA 02] SPARC, The Case for Institutional Repositories: A SPARC Position Paper, Août 2002 <<http://www.arl.org/sparc/IR/ir.html>>.
- [WHI 95] Whitehead A. N., Procès et réalité, Essai de cosmologie, Éditions, Gallimard, Paris 1995.